

A practical approach to unveiling auto-catalytic cleavages mediated by *Mxe* GyrA intein and improving the production of authentic bFGF

Hao Wang¹, Xiuhua H¹, T Sivakumar¹, Cheuk Yin Nelson Lai¹, K L Ng¹, C C Lam¹, W K R Wong^{1*}

¹ Division of Life Science, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

Received: December 06, 2018; Accepted: December 21, 2018; Published: December 27, 2018

*Corresponding author: Division of Life Science, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China. Tel: +852 23587299; Fax: +852 23581552, Email address: bcwkrw@ust.hk

Abstract

A controllable cleavage activity of inteins to release target proteins from affinity tags is an attractive approach for the production of heterologous proteins. Expression of CellBD-GyrA-bFGF (CGF), a fusion precursor comprising the cellulose binding domain (CellBD) of an endoglucanase (Eng) encoded by the *cenA* gene of *Cellulomonas fimi* at the N-terminus, *Mxe* GyrA intein (GyrA) in the middle, and human basic fibroblast growth factor (bFGF) at the C-terminus, resulted in incomplete auto-catalytic cleavages and a partial release of free bFGF in the cytosol of *Escherichia coli*. Amino acid substitutions involving Cys1Ala (C1A) and His197Gln (H197Q) performed simultaneously at both termini of GyrA on CellBD-GyrA-bFGF, thus forming the 49-kDa full-length CGF (H197Q; C1A) mutant precursor (49-kDa MP), were found to exhibit an inhibitory effect on the cleavage detected between CellBD and GyrA, thereby enhancing the reservation of intact 49-kDa MP, which was shown to bind well to Avicel. Sequencing analysis showed that 49-kDa MP underwent complex auto-processing activities involving self-cleavages, proteolysis and fusion of degraded fragments to give rise to a variety of smaller intermediates. One such intermediate, 39-kDa (SI), was also found to be able to bind to Avicel. Free bFGF was effectively cleaved from 49-kDa MP and 39-kDa SI bound on Avicel at pH 8.5 or in the presence of succinic acid. Released bFGF was substantiated to possess the authentic structure and potent bioactivity. The findings may pave way for the development of a facile scalable approach for enhancing production of a heterologous protein.

Keywords: *Mxe* GyrA Intein; Site-Specific Mutagenesis; Cellulose Binding Domain; Avicel; bFGF; In Vitro Cleavages; pH 8.5; Succinic Acid

Introduction

Production of heterologous proteins, e.g., hormone and growth factors that possess a defined mature N-terminal sequence, through recombinant DNA technology is a common research objective. Moreover, the target protein is often required to be produced in an economically manner. A conventional method, irrespective of whether in vivo or in vitro manipulations are involved, relies commonly and indispensably on the use of a protease, which may be a peptidase capable of processing a pre-protein undergoing secretion through the cell membrane [1] or a proteolytic enzyme applied under external conditions [2], to separate the target protein from its fusion partner. However, most

of the conventionally available recombinant strategies may not be conveniently and/or cost-effectively undertaken to fully achieve the mentioned procedures. Thus, in a vast number of examples, derivatives or isoforms of various target proteins, which possess altered properties including lower levels of potency, stability and safety performance, have been incorrectly produced instead.

The discovery, characterization and application of inteins, or protein introns, have revolutionized the production of recombinant proteins which share the same primary structures with their native counterparts. The first intein was discovered in the late 1980s [3]; since then, over 600 putative intein genes have been discovered [4]. The ability of inteins to undergo autocatalytic detachment of themselves from their flanking fusion partners, the N- and C-exteins, has enabled their application to the development of expression platforms that allow the production of target proteins possessing the required primary structures.

Employing an intein, *Saccharomyces cerevisiae* vascular membrane ATPase (VMA), our group has been able to demonstrate, for the first time, the possibility of expressing foreign proteins, with human basic fibroblast growth factor (bFGF) as the experimental model, in their authentic mature forms in *Escherichia coli* [5]. Our findings support the notion that inteins might be exploited to facilitate not only the expression of heterologous proteins as soluble products, but also auto-catalytic cleavages of precursors or/intermediate fusions to form target proteins possessing the required primary structures [5]. Interestingly, these unusual activities are not only observed in the most common recombinant host, *E. coli*, but recently, they have also been demonstrated to take place in another frequently employed host, *Bacillus subtilis*, from which the identification of naturally occurring inteins has been scarcely reported [6].

However, expression of fusions formed between inteins and heterologous proteins may not always yield soluble and properly processed final products in *E. coli*. Early attempts to exploit inteins to mediate protein expression frequently ended up in producing the target proteins as biologically inactive inclusion bodies in the cytoplasm. Making use of a combined protocol involving denaturation and renaturation of protein aggregates comprising

commonly three components aligning with an N-terminal protein tag, a central intein and a C-terminal target protein, followed by adhering the renatured proteins to an affinity adsorbent, of which chitin has been commonly employed to provide the binding capacity, and the target proteins may then be cleaved off the adsorbed fusion precursors by modulating the environmental conditions in the affinity column [7, 8].

Notwithstanding the inducibility of self-cleavage activities of inteins by changing the environmental factors, the exact mechanisms regarding how they exactly operate are not well understood [9]. Nonetheless, the activities have been shown to be affected by the intrinsic properties of inteins and their fusion partners including:

- (1) Biochemical features such as sizes and amino acid compositions of the heterologous proteins,
- (2) The relative position of a target protein as to where (at the N- or C-terminus) it is fused with an intein
- (3) The capability of an intein to promote expression and self-cleavage of an attached foreign 'extein' [5, 10].

Given the high level of uncertainty concerning the exact operations of auto-catalytic processing of inteins, results from scattered research showed that point mutations created in certain conserved protein motifs in inteins might have either an inducible or inhibitory effect on the cleavage performance [11, 12]. Furthermore, reducing agents have been shown to promote auto-dissociation of inteins [9, 11].

To help better understand the mechanistic processes of auto-cleavage activities mediated by inteins in *E. coli*, we adopted a systematic approach to undertake the study. With a newly engineered protein platform comprising the cellulose binding domain (CellBD) of an endoglucanase (Eng) encoded by the *cenA* gene of *Cellulomonas fimi* at the N-terminus [13, 14], *Mxe GyrA* intein (GyrA) in the middle, and human basic fibroblast growth factor (bFGF) at the C-terminus, thus resulting in a

fusion precursor: CellBD-GyrA-bFGF, we explored how mutations created on the N- and C-termini of GyrA might affect auto-catalytic cleavages of bFGF from the precursor and its derivatives in *E. coli*. Our findings, which may not only shed light on how structural modifications might interfere with cleavage activities of inteins, but also on how the introduced approach might be exploited to significantly improve production of authentic bioactive bFGF. Moreover, our findings show also that the cleavages could be complicated by proteolytic processing, an intervention that has not been reported previously.

Material and Methods

Bacterial strain and chemicals

E. coli strain JM101 employed as the host for plasmid manipulations and bFGF expression was previously described [5]. Authentic bFGF derived from our previous work was used to raise antibodies in rabbits and employed as bFGF positive control. All chemicals were purchased from Sigma-Aldrich Corporation (St. Louis, MO, USA) unless otherwise specified.

Construction of expression constructs

Plasmids pWKCGF and its three derivatives: pWKC1A, pWKHQ and pWKCAHQ, were derived from the modifications of construct pWK3R as summarized below. An *EcoRI-XbaI* fragment containing the *LacUV5* expression cassette [1], in which the *egf* and *bfgf* genes were fused to the 5'- and 3'-termini of the coding sequence of the *Sce* VMA intein, respectively, was deleted. A new *EcoRI-XbaI* fragment containing the *gyrA* gene, which was fused to the *cenA* gene encoding the cellulose binding domain (CellBD) at its 5'-terminus and the *bfgf* gene at its 3'-terminus, was synthesized by overlap extension PCR, using oligos P1-P8 as the primers (Table 1), and pWK3R, pTWIN1 (#E6901, New England Biolabs, USA) together with pFC as the template [5, 6]. The PCR product was then ligated with pWK3R restricted with *EcoRI-XbaI* to result in construct pWKCGF (Fig. 1a). To engineer the three pWKCGF derivatives: pWKC1A (C1A), pWKHQ (H197Q) and pWKCAHQ (C1A, H197Q) (Fig. 1b), their inserts were first obtained by site

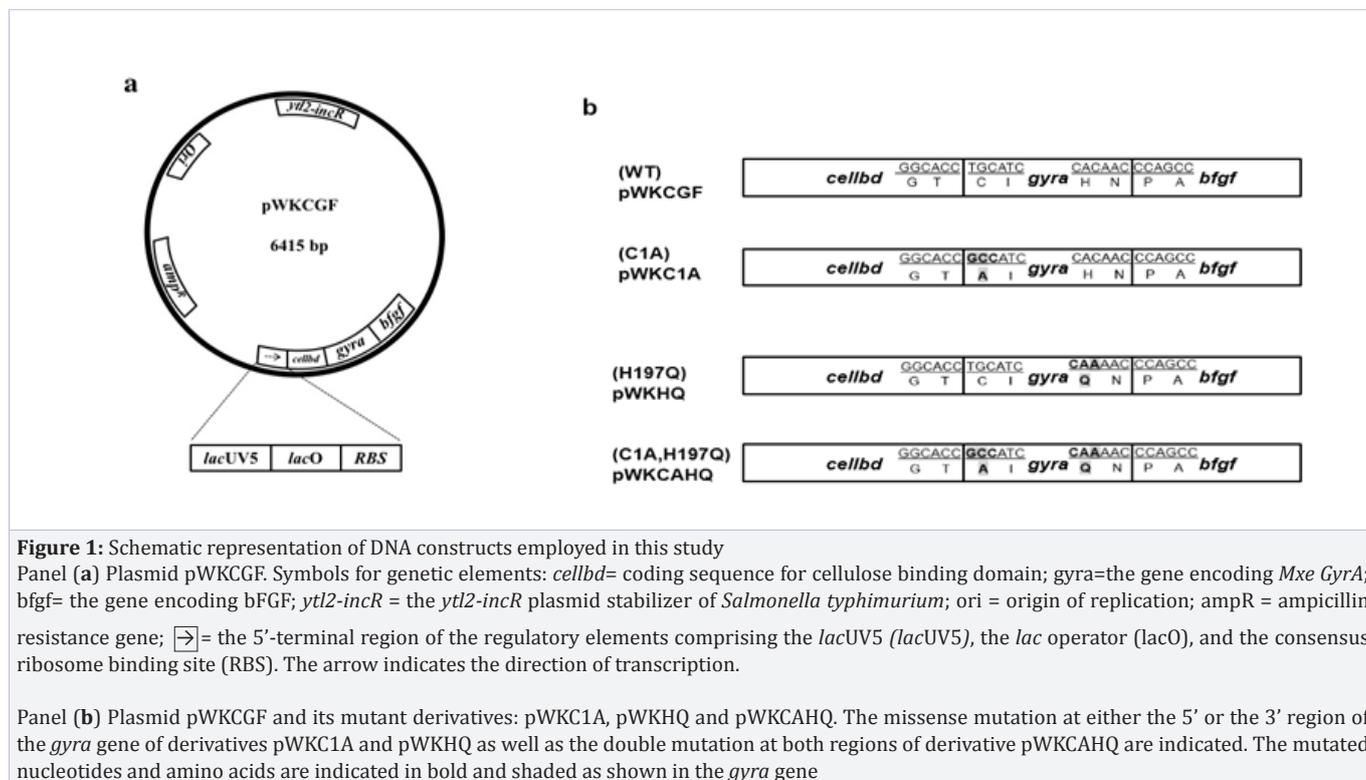
Table 1: Primers employed in this study

Primers	Orientation	Sequence *
P1: <i>EcoRI</i> LacUV5 F	Forward	5'-AGAATTCGCATACTTACTCCCCATCC-3'
P2: <i>LacUV5</i> CellBD F	Forward	5'-CAGGAGGAAAAAATTATGGCTCCCGGCTGCCCGTCGAC-3'
P3: <i>LacUV5</i> CellBD R	Reverse	5'-GTCGACGCGGCAGCCGGGAGCCATAATTTTTCTCTCTG-3'
P4: CellBD-GyrA F	Forward	5'-CACCTGCACGGGCACCTGCATCACGGGAGATGCA-3'
P5: CellBD-GyrA R	Reverse	5'-TGCATCTCCCGTGATGCAGGTGCCCGTGACAGGTG-3'
P6: GyrA-bFGF F	Forward	5'-GGGTTTCGTAGCCACAACCCAGCCTTGCCAGAGGAT-3'
P7: GyrA-bFGF R	Reverse	5'-CCATCCTCTGGCAAGGCTGGGTTGTGGCTGACGAAC-3'
P8: bFGFXbaI R	Reverse	5'-GAATTAATTTCTAGATCATTATTAGCTTTAGCAGACATT-3'
P9: CellBD-GyrA F	Forward	5'-ACCTGCACGGGCACCGCCATCACGGGAGATGCACT-3'
P10: CellBD-GyrA R	Reverse	5'-TAGTGCATCTCCCGTGATGGCGGTGCCCGTGACAGGT-3'
P11: GyrA-bFGF F	Forward	5'-CGAACGGGTTTCGTAGCCAAAACCCAGCCTTGCCAGAGGAT-3'
P12: GyrA-bFGF R	Reverse	5'-TCCTCTGGCAAGGCTGGGTTTGGCTGACGAACCCGTTTCG-3'

* Restriction sites used in the cloning experiments are underlined

directed mutagenesis employing overlap extension PCR, followed by restriction with *EcoRI-XbaI* to generate the DNA inserts, using mutagenic oligos P9-P12 (Table 1) as the primers and pWKCGF as the template. More specifically, oligos P9 and P10 were employed

as the primers for the mutation of C1A, whereas primers P11 and P12 were used for the mutation of H197Q. The inserts were then inserted into vector pWKCGF that had been restricted with *EcoRI-XbaI* to result in the derivatives.



Shake flask cultivations

JM101 *E. coli* transformants were grown at 34 °C in 50 ml of MMBL medium supplemented with 70 µg ml⁻¹ of ampicillin as described previously [5]. In time-course experiments, the growth medium was inoculated with a freshly grown colony and shaken until an A550 value reached 8, followed by an addition of 50 µl of 0.1 M IPTG into the culture. Aliquots of 1 ml cell culture were collected at intervals of 4 h subsequent to induction for various assays. The cell pellets saved were lysed using chemical treatments as described previously [5].

Mechanical cell disruption

The cell pellet obtained from a 50 ml culture was dissolved in 25 ml of 50 mM Tris-HCl, pH 8.0, and the resuspension was disrupted twice using a French Press (Model FPG12800, Stansted Fluid Power Ltd; Harlow, Essex) with a pressure set at 10 kPa. The lysate was centrifuged at 13,500 rpm for 10 min; the supernatant was then kept on ice for use in Avicel binding purification and in vitro cleavage analysis.

In vitro cleavages of CellBD-GyrA-bFGF and derivatives

The application of Avicel, a microcrystalline cellulose powder (Cas No.9004-34-6, Sigma-Aldrich) to the development of a small-scale binding assay was reported previously and the key steps involved are summarized below [15]. To a microfuge

tube containing 0.05 g of Avicel, 1 ml of cell lysate was added and the mixture was incubated on ice for 30 min. Subsequent to centrifugation and removal of the supernatant, 1 ml of 0.5 M NaCl solution was added to the tube. The content was mixed thoroughly, followed by centrifugation and removal of the supernatant. One ml of a 50 mM Tris buffer (at pH 6.0, pH 7.5, pH 8.0, pH 8.5 or pH 9.5) was added and mixed with the pellet, followed by centrifugation and removal of the supernatant. The pellet was then mixed with 25 µl of the 50 mM Tris buffer, and after centrifugation, the supernatant (control sample) was reserved for analysis (experimental control). The pellet was mixed again with 25 µl of the same buffer, and the tube was kept at room temperature for 48 h. The content was then spun down and the supernatant was saved for analysis.

On a larger scale, 5 ml of JM101 [pWKCAHQ] cell lysate was loaded onto a column (Poly-Prep® Chromatography Column, Bio-Rad Pacific Ltd, USA) containing 0.5 g of Avicel. After incubating on ice for 30 min, the column was washed with 40 ml of 0.5 M NaCl solution, followed by washing with 20 ml of a succinic solution (100 mM succinic acid, pH 6.3; 10 mM CaCl₂). The column was then washed twice, each time with 500 µl of the same succinic solution. In both washes, the eluate was reserved for analysis. In the first wash, the eluate (control sample) was collected from the column soon after commencing elution. Regarding the second wash (experimental sample), the buffer was allowed to stay in the column for 48 h at room temperature prior to elution.

Protein analysis and purification

Samples of cell lysates (CL), flow-through fractions (FT), Avicel beads (Avicel), and column eluates were analyzed by Tricine-SDS-PAGE and Western blotting as described previously [5]. Quantification of proteins was performed using the Bio-Rad protein assay according to the manufacturer's instructions (Cas No. #5000002, Bio-Rad Pacific Ltd, USA). Proteins (from 1 ml cell lysate) bound onto (0.05 g) Avicel were detached from the beads by boiling (for 10 min) in an SDS-loading buffer as described previously [15]. The detached proteins were analyzed by Tricine-SDS-PAGE and Western blotting. Images of the Western blot results were quantified by densitometry with the help of the ImageJ software (National Institutes of Health, USA). Resolved protein bands were stained with Coomassie Brilliant Blue and the smaller intermediate (SI) band was retrieved from the gels and subjected to liquid chromatography-tandem mass spectrometry analysis (Thermo Fisher Scientific, CA, USA) as described previously [16].

Collections retrieved from in vitro cleavage treatments were purified by heparin affinity chromatography as describe previously [16], followed by analysis by Tricine-SDS-PAGE. Protein bands were stained with Coomassie Brilliant Blue and candidate bFGF bands were retrieved from the gels and subjected to liquid chromatography-tandem mass spectrometry analysis (Thermo Fisher Scientific, CA, USA) as described previously [16].

Bioassay for bFGF

Experimental samples collected from cleavage experiments were assayed for bFGF activity. The mitogenic effect of bFGF on the proliferation of BALB/c 3T3 fibroblast cells was determined using the MTT assay as described previously [5, 6, 17].

Results

Rationale for the design of the research work

To undertake the study of effects of mutations on auto-catalytic cleavages of inteins, we planned to adopt a systematic approach, with a major goal to obtain reproducible results that would help to fulfill the following two important deliverables: first, enhancing our understanding of operations and controls of auto-processing activities of inteins; second, extending the application potential of inteins, thereby enabling more effective and a wider scope of expression of recombinant proteins. Previous studies on the research topic involved essentially an intein-protein marker fusion, of which the N- or C-terminus of the intein component was fused to an affinity tag, a common example being a chitin-binding domain [7]. Mutations were then created in the intein, and effects of the modifications on the cleavage efficiencies of the protein marker from the intein mutants were assessed. The process commonly involved the expression of the fusion and its derivatives as inclusion bodies, which were then denatured and renatured, followed by binding of the resolubilized proteins to chitin, modulating the environmental conditions to induce cleavages, and lastly, Western blot analysis of the retrieved products [18]. Unfortunately, this approach of study suffers from a major drawback being the presence of misfolded variants in

the solubilized fusion proteins consist of misfolded variants, which might exhibit unpredictable effects other than those of the engineered mutations on the cleavage performance of the fusions.

Therefore, to avoid the complications that might arise from handling protein aggregate, we focused on the application of a soluble fusion precursor comprising an intein, GyrA, and the protein marker, bFGF (Fig. 1), to our study. It was anticipated that the comparison of cleavage activities between wild-type and mutant precursors could be achieved depending essentially on the differences between their primary structures. A facile method exploiting the cellulose-binding domain (CellBD) of an endoglucanase (Eng) encoded by the *cenA* gene of *C. fimi* [13] for use as an affinity tag was devised. Being fused at the N-terminus of GyrA (Fig. 1), CellBD would facilitate the binding of fusion products to Avicel, and bFGF could then be released through in vitro cleavages of the bound proteins for analysis.

Expression of the fusion precursor, CellBD-GyrA-bFGF, in *E. coli*

Western blot analysis of proteins expressed in JM101 cells harboring construct pWKCGF (Fig. 1) revealed that the expected fusion precursor, CellBD-GyrA-bFGF (CGF), was processed quite effectively to form bFGF as the major product (Fig. 2a). Since CGF, its intermediates and free bFGF were detected essentially in the cell lysate (CL) but not in the debris (D), which could be spun down using low-speed centrifugation [19]; both the fused and detached bFGF products were concluded to be existing in soluble forms (Fig. 2d).

The observations from the above study, which included:

- (1) High levels of solubility of the 49-kDa full-length precursor, CGF, and its smaller derivatives (Fig. 2a)
- (2) An efficiently processed CGF precursor to result in free bFGF as a major member (ca. 45%) among the final products
- (3) The presence of a complex of smaller intermediates, despite their lack of identities, which might form a good source for additional yield of bFGF, supported the potential application of CGF to the study of mutational effects on self-cleavage activities and the improvement of bFGF production.

Mutational effects on the processing of CGF

To investigate whether point mutations created at the two junctions:

- (i) Between GyrA and its N-extein, CellBD;
- (ii) Between GyrA and its C-extein, bFGF, might have any negative impact on the cleavage of bFGF from GyrA carried on CGF and its intermediates, three sets of as substitutions were engineered on the peptide sequence of GyrA, which included conversions of His197th to Gln (H197Q), Cys1st to Ala (C1A), and the double mutation comprising both H197Q and C1A (Fig. 1b). The H197Q substitution, involving a substitution of Gln for His on GyrA at the junction of GyrA-bFGF (Fig. 2a), was previously shown to be able to attenuate the C-terminal cleavage activity of GyrA [11]. However, the mutation did not

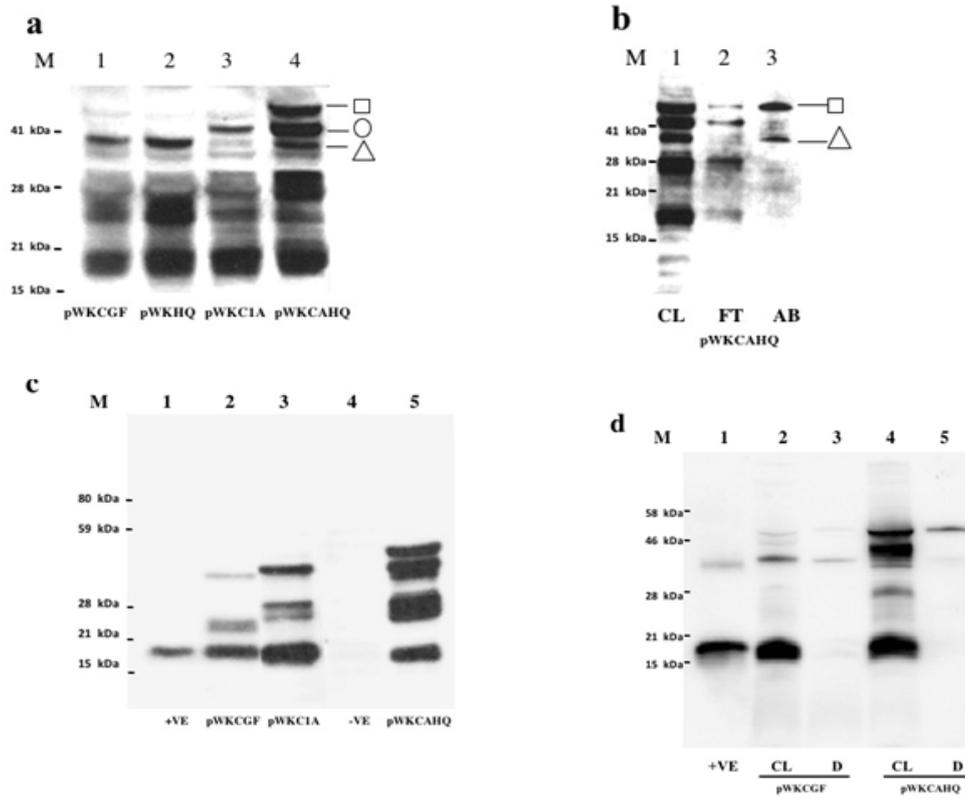


Figure 2: Western blot analysis of proteins presented in the lysates of *E. coli* JM101 transformants carrying various plasmids possessing anti-bFGF activities

Panel (a) Results of the analysis of cell lysates (CL) prepared from transformants harboring the four constructs: *E. coli*, pWKCGF (WT), pWKHQ (H197Q), pWKC1A (C1A) and pWKCAHQ (C1A, H197Q), grown under IPTG induction for 8 h, are shown. Each CL sample loaded was equivalent to 5 μ l of the original cell culture

Panel (b) Results of binding of CL prepared from JM101 [pWKCAHQ] cells to Avicel are shown: lane 1, CL; lane 2, proteins washed off the Avicel column; lane 3, proteins bound to Avicel. The 49- kDa MP precursor (□), 41-kDa (○) and 39- kDa SI (△) intermediate bands are indicated

(c) Results of the analysis of cell lysates (CL) prepared from transformants harboring the three constructs: pWKCGF (WT), pWKC1A (C1A) and pWKCAHQ (C1A, H197Q), grown under IPTG induction for 8 h, are shown. In the analysis, all CL samples contained the same amount of total proteins

(d) CL and debris (D) samples prepared from transformants harboring pWKCGF (WT) and pWKCAHQ (C1A, H197Q), grown under IPTG induction for 8 h, were compared as shown. In the analysis, both CL and D samples contained the same amount of total proteins. Other symbols used include: FT: flow through (unbound proteins); AB: Avicel beads (showing bound proteins); M: protein markers; +VE: bFGF standard [5]; -VE: cell lysate lacking bFGF

appear to work well on improving the overall yields of CGF and intermediates (Fig. 2a).

On the other hand, it was reported that the substitution, C1A, might help block the N-terminal cleavage activity of an intein, thus facilitating the accumulation of a full-length precursor [20]. However, the C1A mutation engineered at the border of CellBD-GyrA (Fig. 2a) appeared to exert neither a serious negative impact on the cleavage of bFGF off the GyrA mutant, C1A, nor a positive effect on improving the yield of the 49-kDa full-length precursor;

despite the presence of a notable improvement in the yield of a smaller, 41-kDa, intermediate (Fig. 2a). Since neither of the two missense mutations, H197Q and C1A, worked well to result in higher yields of the full-length precursor, it was then decided that a new DNA construct, pWKCAHQ (Fig. 1b), which encoded a new CGF mutant, CGF (H197Q; C1A), carrying both the H197Q and C1A substitutions, respectively at the N- and C-termini of GyrA, was engineered. It was hoped that the double mutant would lead to the formation of a higher level of full-length precursor.

Analysis of the detached proteins previously bound to Avicel

The cell lysate of a JM101 [pWKCAHQ] culture was used to bind to Avicel, followed by washing and elution with succinic acid (Material and Methods). Western blot analysis of the eluate revealed the presence of a prominent protein band, which appeared to be the 49-kDa full-length CGF (H197Q; C1A) mutant precursor (Fig. 2a), the protein species expected to be retrieved for use in the development of an in vitro purification and cleavage assay.

Quite unexpectedly, a smaller protein band, with a molecular size of ca. 39-kDa, was revealed to be co-retrieved with the 49-kDa full-length CGF (H197Q; C1A) mutant precursor (49-kDa MP) (Fig. 2b). The significantly smaller 39-kDa intermediate (39-kDa SI), presumably derived from the processing of 49-kDa MP, which was yet capable of adhering to Avicel, was co-eluted with 49-kDa MP (Fig. 2b) by boiling under reducing conditions (Material and Methods).

Sequencing results of 39-kDa SI revealed that this smaller derivative appeared to possess an intact form of bFGF, whereas both CellBD and GyrA had been partially deleted (Fig. 3). The

deletions were postulated to result from proteolytic degradation from both the N- and C-termini of a self-cleavage site, of which the exact location has not been well defined, to result in a deletion mutant comprising a 30-residue CellBD fragment that was fused with a GyrA deletant missing the first 11 aa residues (Fig. 3). Presumably, self-cleavage occurred somewhere at the junction between CellBD and GyrA on 49-kDa MP. The free C- and N- termini of CellBD and GyrA, respectively, then experienced different extents of proteolytic degradation, followed by quite an unexpected event, in which the two processed peptides were connected in tandem as depicted in a schematic diagram of 39-kDa SI (Fig. 3). Moreover, the N-terminus of CellBD was determined to be substantially degraded (Fig. 3). Interestingly, despite possessing only a much shorter polypeptide of 30 residues, which was less than one-third of the full length of CellBD, the CellBD deletant appeared to work well to facilitate the binding of 39-kDa SI to Avicel. This 39-kDa SI was identified when it was co-eluted with 49-kDa MP (Fig. 2b). Nonetheless, despite being a more complex mutation than the C1A substitution initially engineered on 49-kDa MP (Fig. 2b), the deletions of CellBD and GyrA on 39-kDa SI did not appear to have a negative impact on the self-cleavage performance of bFGF from GyrA (see below).

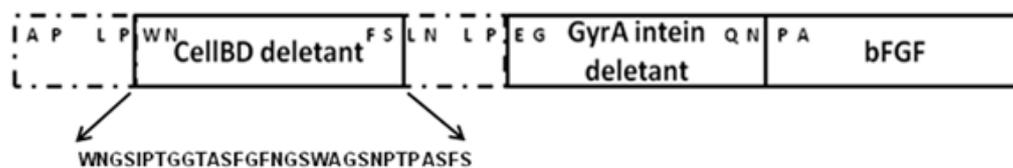


Figure 3: Schematic representation of 39-kDa SI bound to Avicel

The different components of 39-kDa SI, comprising CellBD and GyrA in form of deletion mutants, and bFGF possessing the intact structure. Sequencing results revealed that the CellBD mutant was a 30-residue deletant (missing the N-terminal 67 and the C-terminal 9 aa of the wild-type CellBD) of which the aa sequence is aligned as defined by the two arrow signs. The results also revealed that the GyrA variant was missing the N-terminal 11 aa residues. The aa sequences of the various components in 39-kDa SI are represented by boxes drawn with solid lines, whereas the deleted sequences are represented by boxes sketched with dotted lines. Note that the C-terminal F and S residues on CellBD is fused to the N-terminal E and G residues of GyrA in 39-kDa SI.

In vitro cleavage of full-length and truncated precursors adsorbed on Avicel

To evaluate the effectiveness of bFGF detached from the two precursors: 49-kDa MP and 39-kDa SI, which were bound on Avicel, several buffer systems were tested in the study. It was shown that alkaline environments or addition of succinic acid coupled with CaCl₂ worked effectively to release bFGF from 49-kDa MP and 39-kDa SI, yielding a homogenous product as summarized in (Table 2). A combined weight of 16.60 µg of 49-kDa MP (11.00 µg) and 39-kDa SI (5.60 µg) were estimated to get adsorbed onto 50 mg Avicel. At pH 8.5, a yield of 3.20 µg of bFGF, about 53% of the expected total yield, was estimated to be released from the bound proteins (Table 2). Under other conditions including different alkaline pH values or supplementation of succinic and CaCl₂, bFGF could also be cleaved effectively off the bound proteins comprising 49-kDa MP and 39-kDa SI (Table 2).

Authenticity and mitogenicity of the bFGF product

Notwithstanding the demonstration of in vitro cleavages of the 49-kDa MP and 39-kDa SI mutants to yield free bFGF (Fig. 2), it was important to determine whether the recombinant product and native bFGF share the same biochemical properties. In fact, sequencing and bioassay results substantiated that the bFGF product possessed the 146-residue authentic structure (Table 3) as well as potent bioactivity (Fig. 4). In addition, the sequencing results confirmed also that the bFGF product comprised a homogeneous population of molecules (Fig. 2b), irrespective of their origins from two structurally distinguishable precursors, 49-kDa MP and 39-kDa SI (Fig. 2b).

Table 2: Yields and cleavage efficiencies of bFGF cleaved in vitro from proteins in the lysates of JM101 [pWKCAHQ] cells bound to Avicel under different conditions

	Yield (μg) ^a	Cleavage rate % ^g
50 mM Tris-HCl pH 6.0	not detectable ^b	NA
50 mM Tris-HCl pH 7.5	not detectable	NA
50 mM Tris-HCl pH 8.0	1.60 ± 0.042426 ^c	26.50%
50 mM Tris-HCl pH 8.5	3.20 ± 0.028284 ^d	53.20%
50 mM Tris-HCl pH 9.5	2.40 ± 0.077782 ^e	39.90%
100 mM Succinic Acid (10 mM CaCl ₂ , pH 6.3)	36.30 ± 0.052520 ^f	60.50%

a) Average value based on the calculations from 3 times of repeated experiments

b) Neither SDS-PAGE nor Western blotting resulted in detectable bFGF activities. NA: not available

c, d, e) Yield of bound bFGF resulting from 2 ml of cell culture eluted in 25 μl of elution buffer (see Material and Methods for the details of binding and cleavage)

f) Yield of bound bFGF resulting from 10 ml of cell culture eluted in 500 μl of elution buffer (see Material and Methods for the details of binding and cleavage)

g) The molar ratio between 49-kDa MP and bFGF is 49:16.5; the molar ratio between 39-kDa SI is 39:16.5, and that between the smaller intermediate (39-kDa SI) and bFGF is 49:39:16.5.

Table 3: Analysis of bFGF^a by liquid chromatography tandem mass spectrometry

Peptide ^b	Mr(Calc) ^c	Mr(Expt) ^{d,e}
NH ₂ -PALPEDGGSGAFPPGHFKD	1779.858	1779.488
KAILFLPMSAKS-COOH	1089.626	1088.645

a) Prepared from cleavages of CGF retrieved from the lysate of JM101 [pWKCGF] cells

b) Identification of the N-terminal and C-terminal sequences achieved with the help of the Mascot search engine

c) Theoretical mass-to-charge ratio of the peptide

d) Experimental mass-to-charge ratio of the peptide

e) Virtually the same results obtained from cleavages of 49-kDa MP and 39-kDa SI retrieved from JM101 [pWKCAHQ] cells

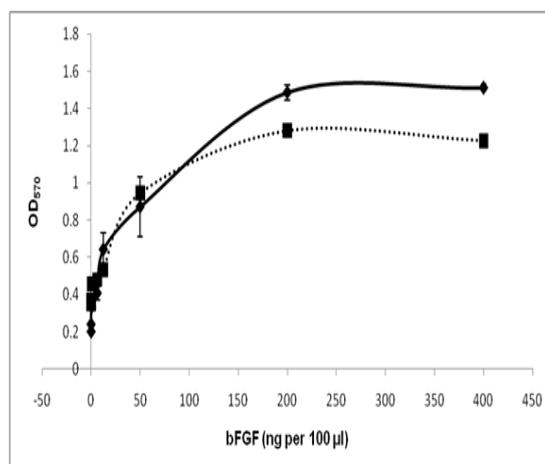


Figure 4: Effects of bFGF samples on promoting the growth of BALB/c 3T3 cells

Cell growth was promoted using different concentrations of bFGF retrieved from in vitro cleavages of 49-kDa MP and 39-kDa SI bound to Avicel (solid line) as described in Fig. 3 and the bFGF standard (dotted line) [17]

Discussion

The introduction of inteins in the late 1980s [3] has revolutionized protein expression employing recombinant DNA approaches. Intein-mediated protein expression has facilitated the implementation of in vitro cleavages of intein-target protein fusions in the absence of auxiliary proteolytic enzymes [7]. Recently, intein-target protein precursors have also been shown to be capable of undergoing auto-catalytic cleavages to result in not only bioactive but also precisely processed authentic final products in the cytoplasm of both *E. coli* [5, 21] and *B. subtilis* [6]. Through genetic enhancement, the in vivo cleavage protocol has been successfully employed to result in high-level expression of authentic bFGF in *E. coli* [5, 17, and 22].

Despite the advantages of protein expression mediated by inteins, our results from auto-catalytic cleavages of the wild-type fusion protein, CellBD-GyrA-bFGF (CGF; Fig. 2a), encoded by construct pWKCGF (Fig. 1a), support the notion that the cleavages might be mechanistically quite complex. Detection of multiple protein bands by Western blot analysis (Fig. 2a) suggested that cleavages of the CGF precursor and smaller intermediates took place first at the junctions between CellBD-GyrA and/or GyrA-bFGF. Moreover, the existence of multiple bands might also imply that the processing event was incomplete, thus resulting in only a partial yield of bFGF, which was estimated to be ca. 45% of the total bFGF expected to be obtained from complete cleavages of the CGF precursor (Fig. 2a). Although incomplete processing of the intermediates might partly explain the presence of multiple bands, it was still unclear why the overall band pattern was notably more complex (Fig. 2a) than it was expected since specific anti-bFGF antibodies were used (Fig. 2c). Therefore, only bFGF and intermediates carrying bFGF were supposed to be positively detected in the assay!

To help understand, first, the biochemical events underlying self-cleavages in precursors formed between inteins and foreign/target proteins, and second, how the yield of a target protein might be improved through better comprehension of the mechanism(s), we started off with the development of a functional assay, which involved the expression of a soluble fusion, CGF, and its derivatives, followed by quantitative and qualitative analyses of the products, using bFGF as the marker, retrieved from various stages of the assay. It was a traceable process, in which CellBD was used as an affinity tag (Fig. 1a) to help retrieve precursors and intermediates that were bound to Avicel for in vitro cleavages. On the other hand, specific anti-bFGF antibodies helped track down candidate molecules and the final product, bFGF (Fig. 2), which were then subjected to in vitro cleavages, followed by authenticity (Table 3) and bioactivity (Fig. 4) analyses.

Effects of mutations involving the substitutions of Cys by Ala (C1A) at the N-terminus and His by Gln (H197Q) at the C-terminus of inteins on cleavage activities of the C-terminal fused proteins were reported previously [20, 23]. Despite inhibitory effects shown by the mutations in these studies, they did not appear to have a significant negative impact on the cleavage of bFGF from its fusion precursors/intermediates (Fig. 2a). However, in

the presence of both C1A and H197Q substitutions, the mutant precursor, 49-kDa MP, expressed by pWKCAHQ (Fig. 1), was revealed to experience an inhibitory effect on the cleavage between CellBD and the GyrA mutant, thereby resulting in a marked increase in the molar quantities of 49-kDa MP and smaller intermediates (Fig. 2a). Interestingly, the yield of bFGF derived from cleavages of 49-kDa MP and the intermediates was not significantly affected, although the amounts of the latter molecules were greatly enhanced (Fig. 2a).

The application of the facile Avicel binding/cleavage protocol (Material and Methods) facilitated not only the adsorption of 49-kDa MP and intermediates harboring CellBD, but also the attainment of both qualitative and quantitative details regarding the bound and unbound molecules. Analysis of the proteins in the lysate of JM101 [pWKCAHQ] cells bound to and subsequently detached from Avicel revealed that 49-kDa MP was one of the two retrieved components (Fig. 2b). Interestingly, the 41-kDa intermediate (Fig. 2a), was revealed not to be another candidate, but instead a smaller (39-kDa) and less prominent polypeptide, 39-kDa SI, was detached from Avicel binding (Fig. 2b). Presumably, through self-processing, the CellBD affinity tag of the 41-kDa intermediate got deleted. However, it was quite surprising to observe that the smaller 39-kDa SI derivative still managed to bind to Avicel. Sequencing analysis unveiled, remarkably, the presence of aa deletions at both ends of CellBD (Fig. 3). It was envisaged that subsequent to auto-cleavage occurring between CellBD and GyrA in 49-kDa MP, CellBD was subjected to proteolytic degradation from both ends to result in a deletant retaining a sequence of only 30 aa (Fig. 3). On the other hand, the GyrA-bFGF fragment encountered an N-terminal proteolytic attack, thereby converting to a deletant missing the first 11 aa (Fig. 3). Quite unexpectedly, the CellBD and GyrA-bFGF deletants were then fused together to form 39-kDa SI. Although exactly how the two shortened fragments approached and joined together is yet to be clarified, the process is unlikely the same as that postulated for protein splicing, in which two initially separated exteins are joined together subsequent to the excision of the central intein. Moreover, further investigation is required to verify exactly where self-cleavage took place at the junction between CellBD and GyrA in 49-kDa MP to result in the mentioned proteolytic deletions.

Irrespective of the variations in size between 49-kDa MP and 39-kDa SI and in the complexity of mutations of the GyrA components at their N-termini (Fig. 1 and Fig. 3), both molecules were not only able to bind to Avicel, but also able to undergo self-cleavage to release bFGF following modifications of the environmental factors. It is noteworthy that increasing the pH value of the cleavage buffer to a range between 8.0 and 9.5 or supplementation of succinic acid to the buffer worked well to trigger self-cleavages of bFGF from the C-termini of both 49-kDa MP and 39-kDa SI (Table 2). It was previously explained that an Asn residue was susceptible to succinimide formation, thus leading to protein degradation [20, 24, and 25]. Therefore, under alkaline conditions (pH 8.0 - 9.5) employed in our work, the more basic environment was expected to facilitate the cleavage of bFGF

from GyrA, of which the C-terminal aa is Asn (Fig. 3). However, the exact function of succinic acid in promoting self-cleavage activities of GyrA is not clear. Previous studies postulated that succinic acid might play a role in protein refolding [26]. Thus, it was possible that the reagent helped the intein fusions achieve the necessary folded structures for self-cleavages.

The susceptibility of CellBD to both N- and C-terminal proteolytic degradation (Fig. 3) might help explain why the band patterns of the proteins present in various lysate samples revealed on Western blots using anti-bFGF antibodies appeared to be so complex (Fig. 2a). Moreover, the analysis revealed that many of the molecules, despite with larger sizes, were unable to bind to Avicel (Fig. 2b), supporting the conclusion that although these molecules contained bFGF, they lacked the ability to interact with Avicel. It would be interesting to see whether the expression of specifically 39-kDa SI (Fig. 3) might result in reduced levels of degraded intermediates (Fig. 2a), thereby helping improve the overall yield of bFGF at two different levels: first, through direct self-cleavages of 39-kDa SI that is expected to be present at a higher concentration due to its resistance to proteolysis; second, through more efficient binding, due to the lack of competition, e.g., from 49-kDa MP, of 39-kDa SI to Avicel. As a result, a higher yield of bFGF is expected to be detached from the bound molecules. Moreover, with less intermediate derivatives, less non-specific interference, the binding of intact 39-kDa SI to Avicel will also be more efficient. Thus, optimum cleaving conditions, it is expected that a homogeneous population of 39-kDa SI would perform significantly better than a mixed population comprising 39-kDa SI and other structural variants, e.g., 49-kDa MP (Table 2) in providing improved yields of bFGF. Since incomplete auto-catalytic cleavages are commonly encountered in intein-mediated protein expression [5, 7], the novel in vitro enrichment/self-cleavage approach introduced in this communication may furnish an effectual process for enhanced production of authentic and bioactive heterologous proteins.

Acknowledgements

This study was funded by RGC project: GRF16101515 and Research Contracts: 13142580CLIL07W011 and 1617219-0. We thank Biosciences Central Research Facility at HKUST and the Instrumental Analysis Center of Shenzhen University (Xili Campus) for their mass spectrometry services.

Compliance with ethical standards

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Sivakesava S, Xu ZN, Chen YH, Hackett J, Huang RC, Lam E, T L Lam, et al. Production of excreted human epidermal growth factor (hEGF) by an efficient recombinant *Escherichia coli* system. *Process Biochem.* 1999;34(9):893–900.
2. Li Y. Self-cleaving fusion tags for recombinant protein production. *Biotechnol Lett.* 2011;33(5):869–881. doi: 10.1007/s10529-011-0533-8
3. Hirata R, Ohsumk Y, Nakano A, Kawasaki H, Suzuki K, Anraku Y. Molecular structure of a gene, VMA1, encoding the catalytic subunit of H (+)-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *J Biol Chem.* 1990;265(12):6726–6733.
4. Shah NH, Muir TW. Inteins: nature's gift to protein chemists. *Chem Sci.* 2014;5(1):446–461.
5. Kwong KWY, Wong WKR. A revolutionary approach facilitating co-expression of authentic human epidermal growth factor and basic fibroblast growth factor in both cytoplasm and culture medium of *Escherichia coli*. *Appl Microbiol Biot.* 2013;97(20):9071–9080. doi: 10.1007/s00253-013-5090-8
6. Hu X, Lai CY, Thiyagarajan S, Wang H, Ng KL, Lam CC, Wong WKR, et al. Novel Strategy for Expression of Authentic and bioactive human basic fibroblast growth factor in *Bacillus subtilis*. *Appl Microbiol Biot.* 2018;102(16):7061–7069. doi: 10.1007/s00253-018-9176-1
7. Zhang Y, Zhang K, Wan Y, Zi J, Wang Y, Wang J, Xue X, et al. A pH-induced, intein-mediated expression and purification of recombinant human epidermal growth factor in *Escherichia coli*. *Biotechnol Progr.* 2015;31(3):758–764.
8. Sun Z, Chen J, Yao H, Liu L, Wang J, Zhang J, Liu JN, et al. Use of Ssp dnaB derived mini-intein as a fusion partner for production of recombinant human brain natriuretic peptide in *Escherichia coli*. *Protein Express Purif.* 2005;43(1):26–32.
9. Chong S, Shao Y, Paulus H, Benner J, Perler FB, Xu MQ. Protein Splicing Involving the *Saccharomyces cerevisiae* VMA Intein: The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system. *J Biol Chem.* 1996;271(36):22159–22168.
10. Marshall CJ, Grosskopf VA, Moehling TJ, Tillotson BJ, Wiepz GJ, Abbott NL, Raines RT, et al. An evolved *Mxe GyrA* intein for enhanced production of fusion proteins. *ACS Chem Biol.* 2014;10(2):527–538. doi: 10.1021/cb500689g
11. Chong S, Montello GE, Zhang A, Cantor EJ, Liao W, Xu MQ, J Benner, et al. Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. *Nucleic Acids Res.* 1998;26(22):5109–5115.
12. Wood DW, Wu W, Belfort G, Derbyshire V, Belfort M. A genetic system yields self-cleaving inteins for bioseparations. *Nature biotechnology.* 1999;17(9):889–892.
13. Wong WKR, Gerhard B, Guo ZM, Kilburn DG, Anthony R, Warren J, Miller RC Jr, et al. Characterization and structure of an endoglucanase gene *cenA* of *Cellulomonas fimi*. *Gene.* 1986;44(2-3):315–324.

14. Damude HG, Ferro V, Withers SG, Warren RA. Substrate specificity of endoglucanase A from *Cellulomonas fimi*: fundamental differences between endoglucanases and exoglucanases from family 6. *Biochem J*. 1996;315(Pt 2):467-472.
15. Meinke A, Gilkes NR, Kilburn DG, Miller RC, Warren RA. Cellulose-binding polypeptides from *Cellulomonas fimi*: endoglucanase D (CenD), a family A beta-1, 4-glucanase. *J Bacteriol*. 1993;175(7):1910-1918.
16. Kwong KWY, Ng KL, Lam CC, Wang YY, Wong WKR. Authentic human basic fibroblast growth factor produced by secretion in *Bacillus subtilis*. *Appl Microbiol Biot*. 2013;97(15):6803-6811. doi: 10.1007/s00253-012-4592-0
17. Kwong KWY, Sivakumar T, Wong WKR. Intein mediated hyperproduction of authentic human basic fibroblast growth factor in *Escherichia coli*. *SCI REP*. 2016.
18. Guo C, Li Z, Shi Y, Xu M, Wise JG, Trommer WE, Yuan J, et al. Intein-mediated fusion expression, high efficient refolding, and one-step purification of gelonin toxin. *Protein Expres Purif*. 2004;37(2):361-367.
19. O'Neill GP, Kilburn DG, Warren RA, Jr MR. Overproduction from a cellulase gene with a high guanosine-plus-cytosine content in *Escherichia coli*. *Appl Environ Microb*. 1986;52(4):737-743.
20. Belfort M, Wood DW, Stoddard BL, Derbyshire V. *Homing endonucleases and inteins*. Springer Berlin. 2013.
21. Kwong KWY, Ng AKL, Wong WKR. Engineering versatile protein expression systems mediated by inteins in *Escherichia coli*. *Appl Microbiol Biot*. 2016;100(1):255-262. doi: 10.1007/s00253-015-6960-z
22. Thiyagarajan S, Lai CY, Lam CC, Hu X, Wang H, Ng KL, Wong WKR, et al. Purification of Authentic Human Basic Fibroblast Growth Factor Expressed in Both the Cytoplasm and Culture Medium of *Escherichia coli*. *Int J Chromatogr Sep Tech: IJCST*. 2018.
23. Chen L, Benner J, Perler FB. Protein splicing in the absence of an intein penultimate histidine. *J Biol Chem*. 2000;275(27):20431-20435.
24. Clarke S. Propensity for spontaneous succinimide formation from aspartyl and asparaginyl residues in cellular proteins. *Int J Pept Protein Res*. 1987;30(6):808-821.
25. Teshima G. Deamidation in Proteins and Peptides. *IonSource.com*. 2000;1-8.
26. Ugwu SO, Apte SP. The effect of buffers on protein conformational stability. *Pharm Tech*. 2004;28(3):86-109.